

유비쿼터스 환경에서의 매장 추천을 위한 추천시스템 개발

김재경^a, 채경희^a

^a 경희대학교 경영대학 경영학과
서울 동대문구 회기동 1번지, 130-701

Tel: +82-2-961-9355, Fax: +82-2-967-0788, E-mail: {jaek, ckh74}@khu.ac.kr

요약

최근 유비쿼터스 환경이 대두됨에 따라 정보의 밀도가 높아지고 있으며, 기업에서는 고객이 제품을 구매함과 동시에 고객의 정보를 저장하여 활용할 수 있게 되었다. 이와 같은 환경은 고객의 요구사항을 사전에 미리 파악하여 적절한 시점과 상황에 맞는 정보를 전달할 수 있도록 하는 추천시스템에 대한 필요성을 증대시켰으며, 다양한 영역에서 추천시스템과 관련된 연구들이 활발하게 이루어지고 있다.

지금까지의 추천시스템은 주로 제품 중심으로 논의되어 왔으나, 유비쿼터스 시장 환경에서는 매장에 대한 논의가 필요하게 되었다. 이는 고객이 다양한 매장을 방문할 수 있으며, 동일한 제품이라도 여러 매장에 동시에 존재할 수 있고, 매장 간의 동선이나 매장의 위치 및 분위기, 제품의 품질이나 가격 등에 대한 개인 선호도에 따라 같은 제품이라도 선호하는 매장은 다를 수 있기 때문이다.

따라서 본 연구에서는 고객의 선호도를 기반으로 유비쿼터스 시장 환경에 적합한 매장 추천시스템을 제안하고자 한다. 매장 추천시스템은 협업 필터링을 기반으로 하고 있으며, Apriori 알고리즘을 이용하여 관련성이 높은 매장들의 집합을 찾아 추천한다. 이 시스템은 기업보다는 고객 중심의 서비스를 제공해 줌으로써 고객의 쇼핑 효율성을 제고시킬 뿐 아니라 장기적인 관점에서 시장 활성화에 기여할 수 있을 것으로 기대한다.

Keywords:

유비쿼터스, 매장 추천시스템, 협업 필터링, Apriori 알고리즘

서론

최근 정보통신 기술의 발전에 따라 유비쿼터스 환경에 대한 논의가 다양한 관점에서 이루어지고 있다. 유비쿼터스 환경은 물리적 공간과 가상

공간의 구분이나 사용에 대한 시간 및 장소 제약 없이 네트워크에 접근하여 정보를 송수신할 수 있는 환경을 의미한다. 이와 같은 유비쿼터스 환경은 편재성, 정보의 풍부성, 상호작용, 높은 정보밀도, 개인화 및 고객화 등의 몇 가지 특징들을 갖는다(Laudon and Laudon, 2007). 그 중 정보밀도는 이용 가능한 정보의 양과 질을 의미하는데, 정보 밀도가 높아짐에 따라 고객의 과거 구매 기록 및 형태에 대한 대량의 정보가 구매와 동시에 저장되고 사용되고 있다. 이를 기반으로 기업에서는 특정 고객에게 맞는 개인화된 정보 전달이 가능하게 되었으며, 고객은 자신들의 성향에 따라 스스로 제품 및 서비스를 바꾸는 고객화도 가능하게 되었다. 특히 개인화된 정보에 대하여 다양한 목적으로 연구가 지속되어오고 있는데, 그 중에서도 고객이 자신의 의도를 명확하게 요구하지 않더라도 고객의 요구사항을 사전에 미리 파악하여 제공하는 추천시스템에 대한 연구가 활발하게 이루어지고 있다(Park, 2003).

추천시스템은 고객의 성향이나 기호에 따라 관심을 가질 만한 상품을 미리 찾아내어 권함으로써 고객의 정보탐색 비용을 줄이고 기업의 판매 기회를 확대시키는 등, 상거래 환경에 긍정적인 영향을 주는 것으로 알려져 있다(Schafer et al., 1999). 추천시스템은 90년대 중반 이후부터 지금까지 여러 가지 방법론과 실험들을 통해 다양한 분야에 적용되어 유용성이 입증되어 왔으며(Rensnick et al., 1994; Hill et al., 1995; Schafer et al., 1999;), 최근에는 전자상거래를 기반으로 고객이 관심을 가질 만한 제품들을 찾아내어 추천하는 방식(Jin et al., 2005)으로 이루어져 왔다. 그러나 지금까지 대부분의 추천시스템 관련 연구는 제품 또는 정보를 추천대상으로 하고 있으며, 고객에게 적합한 매장을 추천하는 연구는 없는 것으로 알려져 있다. 이는 특정 기업이나 상점에서 자신들이 취급하는 상품의 판매를 촉진시키고 해당 상품에 대한 정보를 보다 효과적으로 고객에게 전달하기 위한 목적으로 추천 시스템을 활용해 왔기 때문이다.

기존의 인터넷 공간과는 다르게, 유비쿼터스 환경에서는 고객의 동선에 따라 매장 정보를 제공하는 등의 매장에 대한 관심이 대두되고 있다(Keegan, 2004). 이는 동일한 제품이라도 여러 매장에 동시에 존재할 수 있으며 매장 간의 동선이나 매장의 위치 및 분위기, 제품의 품질이나 가격 등에 대한 개인 선호도에 따라 같은 제품이라도 선호하는 매장은 다를 수 있기 때문에 Krulwich(1997)이나 Keegan(2004)의 연구에서 매장에 대한 논의를 다루고자 하였다.

Krulwich(1997)의 연구에서는 사용자의 인구통계학적인 정보와 전자상거래 상에서 발생한 구매 정보를 이용하여 상품 및 매장을 추천하는 시스템(Lifestyle finder)을 개발하였으며, Fano(1998)의 연구에서는 사용자가 자신의 구매 목록을 전송했을 경우, 위치를 추적하여 현재 위치에서 가장 가까운 매장 정보를 제공해 주는 시스템(Shopper's Eye)을 개발하였다. 그리고 Keegan(2004)의 연구에서는 고객이 전송한 구매 목록 데이터와 고객의 현재 위치를 분석하여 고객에게 가격, 판매 후 관리, 근접성 등의 조건이 좋은 매장의 정보를 제공해주는 시스템(EasiShop)을 개발하였다.

그러나 Krulwich(1997)의 연구는 전자상거래 환경을 기반으로 하여 물리적 공간의 특징을 반영하지 못하고 있으며, Fano(1998)의 연구에서는 물리적 공간의 특징을 반영하기는 하였으나, 매장의 위치 정보만을 사용하여 고객의 선호를 반영하지 못하였다. 그리고 Keegan(2004)의 연구는 제품과 관련된 매장을 찾아내어 해당 매장에 대한 정보를 제공해 줄 뿐, Fano(1998)와 마찬가지로 매장에 대한 고객의 선호나 매장 간의 연관성에 대해서는 고려하지 못하고 있다. 따라서 본 연구에서는 유비쿼터스 환경에 적합하며, 매장에 대한 사용자의 선호 및 매장 간의 연관성을 고려할 수 있는 시스템을 개발하여 고객이 특정 매장에서 제품을 구매했을 경우, 현재의 매장과 관련성이 높은 매장들을 찾아서 추천하고자 한다.

본 연구에서 제시하고자 하는 매장 추천시스템은 많은 연구를 통해 효과성이 입증된 협업 필터링(Collaborative Filtering: CF) 기법(Konstan et al., 1997; Montaner et al., 2003)을 기반으로 Apriori 알고리즘을 혼합한 형태이다. 협업 필터링을 이용한 추천기법은 선호도가 유사한 사용자들의 과거 구매 정보를 기반으로 추천하는 방법이다. 이 때, 선호도가 유사한 고객을 찾아내기 위해 주로 사용되는 방법은 근접 이웃 찾기(nearest neighbor algorithm), 군집분석(clustering), 분류(classification) 등이 있는데(Montaner et al., 2003), 본 연구에서는 자기조직화지도(Self-Organizing Map:SOM)를 이용한 군집분석 방법을 사용하고자 한다. 이는 군집분석 방법의 처리 속도가 근접 이웃 찾거나 분류 방법에 비해 뛰어나며, 특히 자기조직화지도를 이용한 군집분석 방법이 계산 속도나 데이터의 크기와 같은

부분에서 강점을 가지고 있기 때문이다(Kohonen, 1990). Apriori 알고리즘은 추천할 목록을 생성하는 단계에서 사용하게 되는데, 이는 고객의 구매 행동이 발생하기 전에 미리 추천 목록을 생성하여 저장해 놓고, 구매 행동이 발생하면 실시간으로 추천하기 위한 것으로써 추천시스템의 적시성을 높일 수 있다.

본 연구에서는 협업 필터링에 Apriori 알고리즘을 혼합하여 유비쿼터스 환경에서 매장을 추천할 수 있는 시스템을 제시하였다. 이 시스템은 기업 중심이 아닌 고객 중심의 서비스를 제공하기 위한 것으로써, 고객이 선호할 만한 매장을 추천하여 고객의 쇼핑 효율성을 제고시킬 뿐 아니라 장기적인 관점에서 시장 활성화에 기여할 수 있을 것으로 기대한다.

관련연구

협업 필터링 기반의 추천시스템

추천에 대한 문제가 학술적으로 처음 발표된 것은 90년대 중반으로(Rensnick et al. 1994; Hill et al. 1995) 정보검색(IR) 및 데이터 마이닝 분야의 여러 방법을 기반으로 다양한 추천방법들이 제안되어 왔다. 추천시스템을 통해 개인은 정보 탐색 비용을 낮출 수 있고, 기업은 전략적 우위를 확보할 수 있다는 측면에서 중요성이 부각되었고 지속적인 발전을 거듭해오고 있다.

최근 대두되고 있는 유비쿼터스 환경에서는 기기 중심의 컴퓨터 기술이 아닌 사용자 중심의 컴퓨터 기술을 지향하고 있으며, 이와 같은 사용자 중심의 컴퓨터 기술의 핵심에는 추천시스템이 자리잡고 있다. 즉, 고객이 언제 어디서나 편리하게 컴퓨팅 및 네트워크에 접속하여 그들이 원하는 정보를 신속하고 정확하게 전달 받을 수 있도록 하는 것이 추천시스템의 역할인 것이다.

협업 필터링 기법을 이용한 추천 방법, 즉 협업 추천시스템은 지금까지 논의되어온 추천 방법 중에서 가장 성공적인 것으로 알려져 있다(Sarwar et al., 2000). 협업 추천시스템은 이용자와 유사한 사용자들, 즉 이웃(neighbor)들의 선호에 기반한 추천 방법으로써, 추천을 받을 사용자와 가장 유사한 선호 정보를 보이는 이웃들을 찾아낸 후, 이웃들의 구매 정보와 추천 받을 사용자의 정보를 비교하여 이웃들은 가지고 있으나 사용자는 가지고 있지 않는 구매 정보를 찾아서 추천하는 방법이다. 협업 추천시스템은 다양한 형태의 정보에 적용 가능하다는 것과 데이터가 충분할 경우 예측력이 높아진다는 것이 장점으로 분석되고 있으나 정확한 예측을 위해 많은 데이터를 필요로 한다는 것과 이용자 및 콘텐츠 규모가 클수록 많은 계산량을 요구한다는 것, 그리고 정보가 적은 초기 고객들은 정확한 추천이 어렵다는 것과 데이터 희소성 문제

등은 단점으로 파악되고 있다(Konstan et al., 1997). 이와 같은 단점을 극복하기 위하여 최근에는 협업 필터링과 내용기반 필터링을 혼합한 추천시스템들을 통해 초기 사용자 문제, 데이터 희소성 문제 등을 어느 정도 극복하였다(Li et al., 2005). 또한 협업 필터링에 군집분석이나 연관성 분석과 같은 데이터마이닝 기법을 혼합하여 분석의 효율성을 높이기 위한 연구들도 진행되고 있다(Comner and Herlocker, 1999; Lin et al., 2002; Suryavanshi et al., 2005; Xue et al., 2005).

자기조직화지도(SOM)를 이용한 군집분석

선호도가 유사한 이웃을 정의하기 위하여 본 연구에서는 자기조직화지도를 이용한 군집분석을 활용하였다. 즉 같은 군집에 속한 고객들을 하나의 이웃으로 정의하는 것이다. 군집분석은 다양하게 표현된 각 개체들의 변수를 분석하여 동질적인 여러 개의 하위 집단으로 나누는 방법이다. 예를 들어 고객의 나이, 소득 수준 등과 같은 인구통계학적 정보와 구매 금액, 빈도 등과 같은 소비 특성을 기반으로 각 고객들을 분석한 후 유사한 인구통계학적 정보와 소비 특성을 보이는 고객들은 누구인지 분석하는 것이다. 이러한 군집분석에는 유클리드에 의한 거리 계산 방법(Suryavanshi et al., 2005; Xue et al., 2005)과 인공신경망을 학습 방법을 이용한 자기조직화지도 (Kohonen, 1990)가 있다.

자기조직화지도는 Kohonen(1990)에 의해 제시, 개발되어 Kohonen Maps이라고도 알려져 있다. 자기조직화지도 이해하기 어려운 고차원의 데이터를 2차원과 같이 이해하기 쉬운 저차원의 지도(map)로 형상화한다. 이렇게 형상화된 지도는 이해하기 쉽다는 장점 이외에도 입력된 고차원 변수의 위치 관계 그대로 보존된다는 특징이 있다. 다시 말해 실제 공간의 입력 변수가 서로 유사하면 지도 상에서도 가까운 위치에 표현 된다는 것이다. 이러한 자기조직화지도의 특징은 군집 내의 유사성뿐 아니라 군집 간의 유사성도 시각적으로 쉽게 판단할 수 있도록 한다.

본 연구에서는 이러한 자기조직화지도의 위치 보존 특성을 활용하여 이웃이 지나치게 적거나 또는 없는 경우, 즉 같은 군집에 속한 고객의 수가 적거나 없는 경우에는 정확한 추천이 이루어지기 어렵기 때문에 인근 군집과 통합하여 적정 수준의 이웃 수를 유지할 수 있도록 한다.

SOM 모델은 입력층과 출력층, 두 개의 인공신경망 층으로 구성되어 있으며, 입력층은 입력되는 개체의 수와 동일하게 뉴런 수가 존재한다. 그리고 출력층은 사용자가 미리 정해놓은 군집의 수만큼 뉴런 수가 존재하게 되며, 입력층의 개체들은 학습을 통하여 개체간의 거리, 즉 유사성에 의해 출력층에 정렬되는데 이를 지도(map)라 한다. 즉 분류하고자 하는 군집의 수를 정의한 후, SOM을 실시하면 여러 차원으로 표현된

각각의 고객들이 지도상에 정의된 군집으로 분류되는 것이다.

SOM은 반복적인 학습을 통해 입력 벡터와 보다 유사하게 출력층이 형성됨으로써 실제 입력 벡터의 이웃은 지도 상에서도 위상학적 이웃(topological neighbors)으로 유사하게 나타나게 된다. 즉, 다차원의 데이터가 2차원의 지도를 통해 군집으로 표현되고 군집간의 상태 및 거리, 궤도와 같은 것들이 이해하기 쉽게 나타남으로써 프로세스 모니터링, 패턴발견, 오류진단과 같은 다양한 기술 분야에서 활용되고 있다.

Apriori 알고리즘

Breese et al.(1998)의 연구에서는 협업 필터링 기법을 메모리 기반 알고리즘(Memory-based algorithms)과 모델 기반 알고리즘(Model-based algorithms)으로 분류하였다. 전자는 메모리에 의존하는 것으로서 예측을 위해 전체 데이터베이스를 모두 검색하는 것을 의미하고, 후자는 군집분석(clustering), 연관성 분석(association rules mining), 순차적 패턴 발견(sequence pattern discovery) 등과 같은 데이터마이닝 기법들을 이용하여 모델을 구축한 후, 모델의 결과를 이용하여 예측하는 것을 의미한다. 메모리 기반 알고리즘에 비하여 모델 기반 알고리즘은 모델을 구축하는데 추가적인 노력이 소요되기는 하지만, 실제 운영 과정에서는 수행시간이 적게 소요된다는 장점이 있다. 따라서 본 연구에서는 유비쿼터스 환경에서 고객에게 현재의 상황에 맞는 추천 정보를 실시간으로 전달하기 위하여, 후자의 관점에서 협업 추천을 실시하게 되었으며, 이 때 사용되는 알고리즘이 Apriori 알고리즘이다.

Apriori알고리즘은 연관성 분석에 일반적으로 사용되고 있는 알고리즘이다. 연관성 분석은 변수와 변수간의 연관성을 찾아 규칙(rule)의 형태로 나타내는 방법으로 고객이 소매점에서 어떤 제품들을 동시에 구매했는지 살피는 장바구니 분석(market basket analysis)이 대표적이다. 장바구니 분석은 ‘기저귀->맥주’, 즉 “기저귀를 산 사람은 맥주도 동시에 구매한다”와 같이 규칙을 통해 제품간의 관계를 나타내는 것이다. 이 때, 두 제품을 규칙으로 표현하기 전에 {기저귀, 맥주}와 같이 관련성 높은 제품의 집합, 즉 빈발 항목집합(frequent itemset)을 발견하는 방법이 Apriori알고리즘이다. 빈발 항목집합은 최소 지지도 이상을 만족하는 제품들의 집합을 의미하며, Apriori알고리즘의 경우 한 개의 제품만을 포함하는 1-빈발 항목집합에서부터 시작하여 집합에 포함되는 제품의 개수를 하나씩 늘려가면서 최소 지지도 이상을 만족하는 가장 큰 집합(most frequent itemset)을 찾아나간다.

Apriori알고리즘 대해 구체적으로 살펴보면 그림 1과 같다. 이와 같은 Apriori알고리즘은 공집합을

제외한 k 번째 빈발 항목집합의 부분집합은 이전 빈발 항목집합, 즉 $k-1$ 번째 빈발 항목집합에 반드시 포함되어 있어야 하는 속성을 지니고 있다(Han, 2001).

본 연구에서는 서로 관련성이 높은 매장들의 집합을 발견하기 위하여 Apriori 알고리즘을 사용하며, 그림 1과 같은 과정을 통해 생성된 매장의 빈발 집합들(L_k)은 고객 구매 행동에 따른 실시간 추천이 가능하도록 사전에 생성되어 데이터베이스에 저장된다. Apriori 외에도 FP-growth, H-mine, OP, Inverted Matrix 등 다양한 알고리즘들이 관련성이 높은 집합들을 발견하기 위해 사용되고 있다(Agrawal and Srikant, 1994).

C_k : 크기 k 의 후보항목집합
 L_k : 크기 k 의 빈발항목집합

1. 모든 트랜잭션에서 각 아이템의 발생 빈도를 계산하여 C_1 을 검색한다.
2. C_1 중에서 최소지지도 이상의 항목집합을 찾아 L_1 을 생성한다.
3. 다음으로 L_1 을 조합($L_1 \times L_1$)하여 L_2 를 결정하기 위한 C_2 를 생성한다.
4. 더 이상의 L_k 를 생성할 수 없을 때까지 위의 과정을 반복한다. 단, L_k 의 부분집합은 L_{k-1} 에서도 빈발 항목집합이어야 한다. 또한 L_k 는 최대 빈발항목집합이라고 한다.

그림 1 - Apriori 알고리즘에서 빈발 항목집합을 찾아가는 과정

유비쿼터스 환경에서의 매장 추천시스템

본 연구에서 제안하는 추천시스템은 제품이 아닌 매장을 추천하는 시스템으로써 매장에 대한 고객들의 선호 요소, 즉 가격, 친절, 위치 등과 같은 요소가 존재하며, 이와 같은 선호 요소에 따라 특정 고객층이 주로 방문하는 매장들 간에는 연관성이 존재한다는 것을 전제로 하고 있다. 또한 고객들의 선호하는 요소는 연령대나 수입, 지출 규모, 쇼핑 유형에 따라 다를 수 있다. 예를 들어 나이가 많은 고객들이 여러 매장을 방문할 때는 동선이 짧은 매장들을 선호할 수 있으며, 수입이 많고 지출의 규모가 큰 고객들에게 있어 가격은 매장에 대한 선호도를 결정하는 요소가 아닐 수도 있다. 따라서 본 연구에서는 고객 특성에 따른 매장의 선호도 차이를 분석하고 유사 특성의 고객들이 주로 제품을 구매하는 매장들에 대한 관련성을 파악하여 추천하는 시스템을 개발하고자 하며, 이를 위해

고객의 인구통계학적 데이터와 구매 관련 데이터를 분석한다

본 연구에서 제시하고자 하는 매장 추천시스템은 크게 두 단계로 나누어 볼 수 있는데, 첫 번째는 개인이 특정 구매영역에 들어와서 추천 받고자 하는 의사를 표현했을 때 매장을 추천하는 단계이며, 두 번째는 실제 고객이 매장에 방문하여 제품을 구매하였을 때 매장을 추천하는 단계이다. 이때, 특정 구매영역은 고객 및 고객의 구매 정보가 추적되어 하나의 서버에 누적되고 관리되는 영역을 의미하며, 추천 받고자 하는 의사는 고객이 작성한 구매 목록을 특정 구매영역의 서버에 전송하는 것으로써 파악할 수 있다. 그리고 나서 고객이 특정 매장에 방문하여 제품을 구매하면, 결제 등을 통하여 제품 구매가 확인되는 즉시, 그 매장과 관련성이 높은 매장들을 추천하게 된다.

추천시스템의 전체 프로세스에 대해 설명하면 그림 2와 같다. 먼저 고객 군집 모듈에서 고객 프로파일과 구매정보를 이용하여 고객들을 군집으로 분류하여 정의한 후, 선호정보 분석 모듈에서 각 군집별로 두 종류의 선호정보를 분석하여 데이터베이스에 저장한다. 이 두 개의 모듈은 매주, 또는 매월 임의로 선택된 기간마다 실행되어 군집별 선호정보를 갱신한다.

그리고 나서 고객의 행동에 따라 구매목록 기반 추천 모듈과 Apriori 기반 추천 모듈이 실시간 실행됨으로써 추천이 이루어진다. 고객이 특정 구매영역에 들어와서 구매 목록을 전송하면, 구매목록 기반 추천 모듈에서 해당 고객이 어느 군집에 포함되어 있는지 확인한 후, 소속 군집에 대한 선호정보를 검색하여 구매 목록에 포함된 각 제품에 대하여 동일 군집의 고객들이 가장 빈번하게 방문한 매장을 추천한다. 그리고 고객이 추천리스트를 받아 살펴본 후, 매장을 선택하여 제품을 구매하고 나면 Apriori 추천 모듈에서 해당 매장과 관련된 매장들 중 가장 빈번하게 발생하는 매장들을 찾아 집합의 형태로 구성하여 추천한다. 이 때, 현재 제품을 구매한 매장, 즉 추천 직전 시점의 매장 정보만을 이용하여 매장 집합을 구성하고 추천하고자 한다.

다음으로 각 데이터베이스들에 대해 살펴보면 다음과 같다. 고객 프로파일 데이터베이스에는 고객의 최초 거래시 입력되는 인구통계학적 정보와 고객의 군집 분석 실시 후 정의되는 고객의 군집 정보가 저장되어 있다. 구매 정보 데이터베이스에는 고객이 구매한 제품의 이름, 금액, 수량, 제품을 판매한 매장 등의 각 고객별 구매 관련 정보가 저장되어 있으며, 이는 고객이 제품을 구매한 즉시 고객 결제정보를 통해 데이터가 추가된다. 그리고 고객 군집 후에 생성되는 군집별 선호정보는 제품별 매장정보 데이터베이스와 빈발 집합 데이터베이스로 나뉜다.

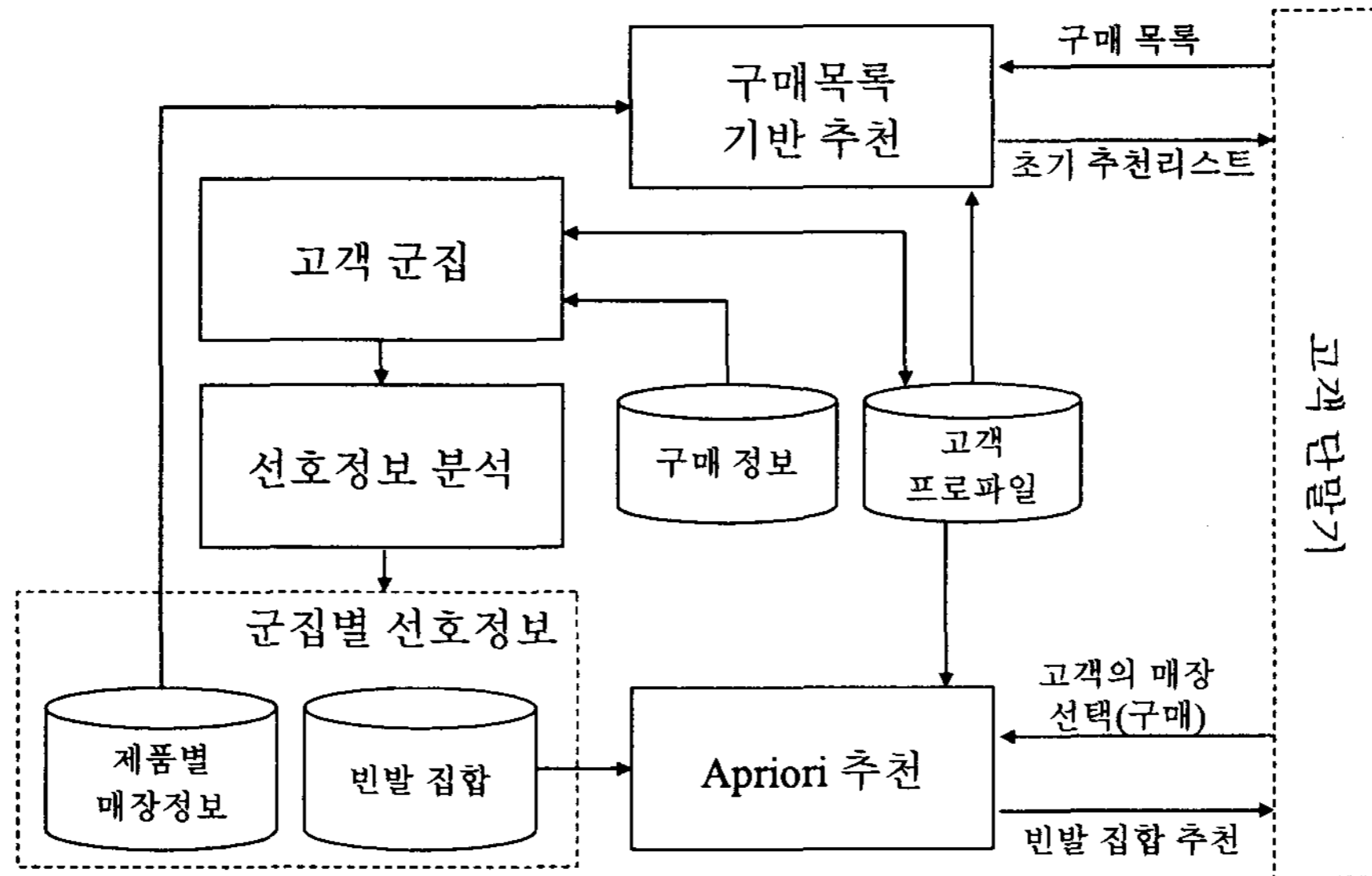


그림 2- 매장 추천시스템 아키텍처

제품별 매장정보 데이터베이스에는 추천영역에 포함된 모든 매장들과 매장에서 판매하고 있는 모든 제품들이 정렬되어 있으며, 고객 군집모듈에 의해 정의된 각 고객 군집별로 제품에 대한 선호 매장들의 순위가 정의되어 있다. 또 다른 선호정보로서 빈발 집합 데이터베이스에는 Apriori 알고리즘을 이용하여 각 군집별로 생성한 매장들의 집합이 저장되어 있는데, 고객의 구매가 발생하면 Apriori 추천 모듈이 빈발 집합 데이터베이스로부터 상황에 맞는 매장 집합을 찾아내어 추천하게 된다. 각 모듈과 연동되는 데이터베이스에 대한 자세한 설명은 다음과 같다.

고객 군집 모듈

고객 군집화 모듈에서는 고객들에 대한 인구통계학적 특성, 구매 활동에 대한 특성들을 고려하여 각각의 고객들을 군집으로 분류하며, SOM을 이용한 군집 분석은 SPSS사에서 제공하는 클레멘타인 5.1을 사용하였다. 군집분석을 위해 고객 프로필과 구매 정보로부터 수집하게 될 데이터를 표 1과 같다.

표 1- 고객 군집에 사용되는 데이터 예시

고객ID	고객 프로필			구매 정보(연간/총)		
	나이	성별	소득 수준	방문 횟수	구매 금액	구매 량
C0001	34	여	중	4	120	19
C0002	30	여	중	9	150	10

각 고객들은 SOM을 통해 군집으로 분류되어 그림 3과 같이 나타나며, 그림 3의 예는 7x5, 총 35개 군집으로 고객이 분류된 것이다. 각 군집은 x축과 y축에 의해 C_{xy} 로 정의 될 수 있으며, 이와 같이

분류된 군집은 위치보존적 특성에 따라 군집 내에 포함된 고객들의 특성이 실제로도 유사한 것은 물론 $A(C_{43})$, $B(C_{53})$ 같이 가까운 군집 간에도 유사성이 존재한다. 따라서 B군집에 속한 고객의 수가 임계치 이하로 적어서 분석이 어려운 경우 A 군집에 포함하여 분석할 수 있다.

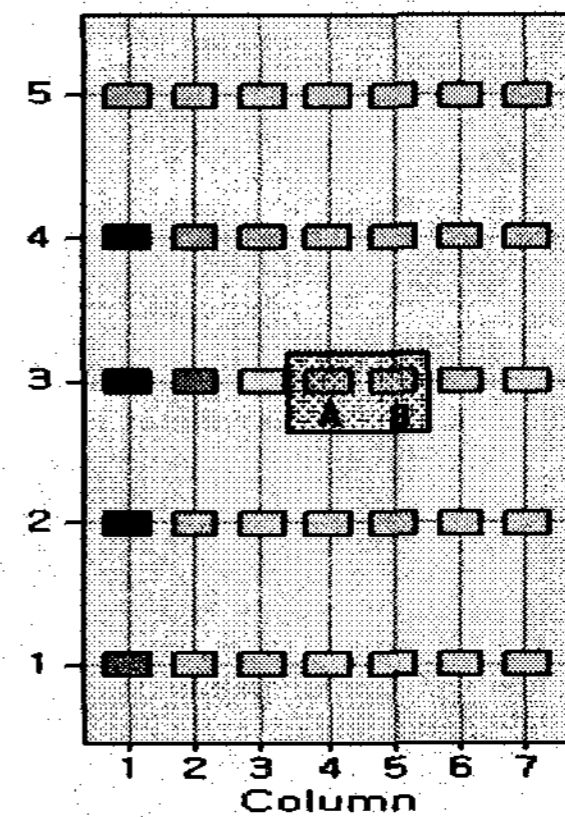


그림 3- SOM 학습을 통해 결정된 고객 군집의 예

관련 연구의 자기조직화지도를 이용한 군집분석을 통해 이미 설명한 바와 같이 SOM은 분류하기에 앞서 적절한 군집 수를 결정해야 한다. 군집의 수를 결정할 때, 군집의 수가 지나치게 많으면 과도적합되어 해당 군집에서 도출된 분석 결과의 일반화가 어렵고 군집의 수가 지나치게 적으면 과소적합되어 데이터의 특성을 제대로 반영하기 어렵다(Han and Kamber, 2001). 따라서 정의된 군집 수가 적절한 판단하는데 주로 사용되고 있는 Davies and Bouldin Index(DB Index)를 이용하여 군집 수의 적절성을 판단하고자 한다(Davies and Bouldin, 1979). DB Index는 분류된 군집들을 서로 비교하여 군집 내의 거리는 가깝고 군집 간의 거리는 멀수록 적절한 분류라고

판단하는 방법이며, K개의 군집에 대한 DB Index를 살펴보면 식 (1)과 같다. 식 (1)에서 S_k 는 k번째 군집의 중심과 그에 속한 개체들 간의 거리 평균을 의미하며, d_{ik} 는 i번째 군집의 중심과 k번째 군집의 중심 간의 거리를 의미한다. 따라서 분류된 군집에 대해 DB(K) 값이 낮게 나타날수록 적절하게 분류된 것으로 판단할 수 있다.

$$DB(K) = \left(\frac{1}{K} \right) \sum_{k=1}^K \max_{k \neq i} \left\{ \frac{S_k + S_i}{d_{ik}} \right\} \quad (1)$$

선호정보 분석 모듈

고객들이 적절한 군집으로 분류되고 나면, 선호정보 분석 모듈에서 각 군집에 대하여 제품별 매장 선호도와 빈발 집합을 분석하여 데이터베이스에 저장한다. 먼저 제품별 매장 선호도에 대해 살펴보면 이 정보는 구매목록 기반 추천 모듈에 사용되는 것으로써 각 군집에서 구매한 제품들과 제품들에 대한 매장의 비율을 정리해 놓은 것이다. 예를 들어 그림 3의 C_{11} 군집에 속한 고객들의 제품별 매장 선호도를 살펴본다면 표 2와 같은 것이다. 구매한 제품은 구매정보 데이터베이스를 검색하여 각 군집별로 분류해 놓은 것이며, 선호 매장은 그 군집에서 해당 제품에 대해 어느 매장을 주로 방문했는지에 대한 비율을 나타낸다.

표 2- 각 군집에 대한 제품별 매장 선호도

군집ID	제품	선호 매장(비율)
C_{11}	신발	S1(0.4),S3(0.3),S4(0.1)...
	청바지	S1(0.1),S2(0.3),S4(0.2)...
	배추	S6(0.1),S7(0.6)

다음으로 빈발 집합은 Apriori 추천 모듈에 사용되는 정보로써 Apriori 알고리즘을 통해 각 군집별로 생성된 매장들의 집합이 저장되어 있는 것이다. 후보 집합 생성에서부터 빈발 집합의 결정까지 자세한 과정은 Han(2001)에서 살펴볼 수 있으므로 데이터베이스에 저장되는 내용만 살펴보면 표 4와 같다. 예를 들어 그림 3의 C_{11} 에 대한 거래 데이터가 표 3과 같고 최소 지지도 값이 2(22%)라면, 표 4와 같이 빈발 집합이 결정될 것이며, 이와 같은 테이블이 데이터베이스에 저장될 것이다. 이때, 거래 데이터는 한 명의 고객이 marketplace에 들어와서 동시에 제품을 구매한 매장들의 목록을 나열한 것이다.

표 3- 거래 데이터의 예

거래ID	매장 목록
T100	S1, S2, S5
T200	S2, S4
T300	S2, S3
T400	S1, S2, S4
T500	S1, S3
T600	S2, S3
T700	S1, S3
T800	S1, S2, S3, S5
T900	S1, S2, S3

표 4에 나타나 있듯이 첫 번째 후보 집합에서는 S1부터 S5까지 모든 매장이 최소 지지도 이상으로써 제거되는 매장 없이 첫 번째 빈발 집합으로 결정되었다. 그러나 두 번째 후보 집합에서는 {S1,S4}, {S3,S4}, {S3,S5}, {S4,S5}가 부분집합 또한 빈발 집합이어야 한다는 Apriori 속성은 만족하지만, 지지도 값이 각각 '1', '0', '1', '0'으로 최소 지지도 값을 만족시키지 못하여 탈락되고 6개의 집합만이 두 번째 빈발 집합으로 결정되었다. 같은 방법으로 Apriori 속성을 만족하는 세 번째 후보 집합을 생성해 보면 {S1,S2,S3}, {S1,S2,S5}가 되는데, 이는 모두 최소 지지도 이상이므로 세 번째 빈발 집합이 되었으며, 그 이상 Apriori 속성을 만족하는 후보 집합이 없으므로 세 번째 빈발 집합이 최대 빈발 집합(most frequent set)이 된다.

표 4- 생성된 빈발 집합의 예(: 지지도)

군집ID	1-빈발집합	2-빈발집합	3-빈발집합
C_{11}	{S1} :6	{S1,S2} :4	{S1,S2,S3} :2 {S1,S2,S5} :2
	{S2} :7	{S1,S3} :4	
	{S3} :6	{S1,S5} :2	
	{S4} :2	{S2,S3} :4	
	{S5} :2	{S2,S4} :2	
		{S2,S5} :2	

구매목록 기반 추천 모듈

이전까지의 모듈들은 의사결정자의 전략적 목적이나 지식에 의해 주단위, 혹은 월단위와 같이 주기적으로 갱신, 수정되는 것이었으나 지금부터 설명하게 되는 두 가지 추천 모듈은 고객의 행동 따라 실시간으로 실행되는 모듈이다.

고객이 특정 구매영역에 들어와서 구매목록을 전송하면 해당 영역을 담당하고 있는 서버에서 그림 4와 같은 추천 프로세스를 진행시키는데, 이를 구매목록 기반 추천 모듈이라 한다. 구매목록 기반 추천 모듈은 고객이 구매목록을 전송하면 고객 프로파일로부터 구매목록을 전송한 고객의 군집을 확인한 후, 해당 군집의 제품별 매장 선호도를 검색한다. 만일 고객의 목록에 포함된 각 제품별로

해당 군집에서 선호하는 매장이 있다면 각각 상위 N개를 선택하여 추천한다. 만일 목록 중에 선호 매장이 검색되지 않는 제품이 존재한다면, 고객이 속한 군집에서 가장 가까운 이웃 군집을 검색하여 선호도가 높은 상위 N개의 매장을 선택하여 추천하는데, 이와 같은 방법은 적절한 매장을 찾을 때까지 반복된다.

예를 들어 표 2와 같이 C_{11} 에 속한 고객이 구매목록으로 신발과 청바지를 전송했다고 가정해보자. 그러면 구매목록 기반 추천모듈은 C_{11} 에 포함된 고객들의 제품별 매장 선호도 데이터베이스로부터 각 제품에 대해 선호하는 매장 정보를 검색하여 상위 N개를 선택한다. 이때 N의 값이 2라고 가정하면, 고객의 단말기에는 신발에 대해서는 S1과 S3이, 그리고 청바지에 대해서는 S2와 S4가 추천 결과로 나타나게 될 것이다. 이와 같이 초기 추천 목록이 전송되고 난 후, 고객이 구매 행동을 시작하면 Apriori 추천 모듈이 진행된다.

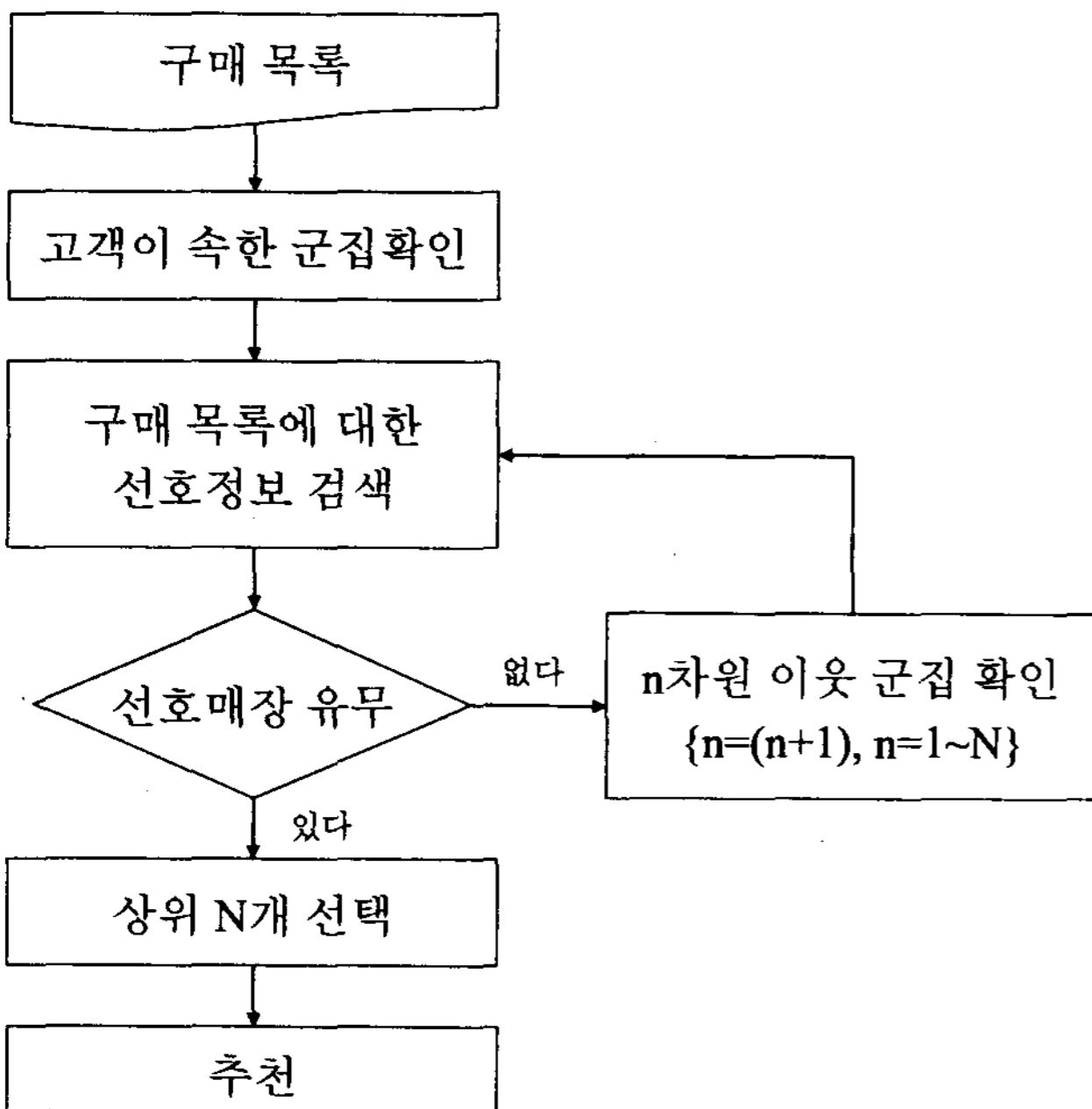


그림 4 - 구매목록 기반 추천 모듈 프로세스

3.4 Apriori 추천 모듈

Apriori 추천 모듈에서는 지속적으로 고객의 구매활동을 탐색하여 구매가 일어나는 순간 해당 매장과 관련된 매장들을 찾아 추천해 주게 되며 자세한 프로세스는 그림 5와 같다. 고객이 제품을 선택하고 결제하면, 고객의 구매정보가 데이터베이스에 저장되는 동시에 빈발 집합 데이터베이스로부터 해당 매장을 포함하고 있는 빈발 집합을 검색한다. 즉 고객과 유사한 고객들의 과거 구매기록을 이용하여 현재 고객이 제품을

구매한 매장과 동시에 방문할 가능성이 높은 매장들을 찾아내는 것으로써, 서로 동시에 방문할 가능성이 높은 매장들을 빈발 집합이라 한다. 이때, 가장 큰 빈발 집합, 즉 k-빈발 집합을 우선적으로 검색하여 추천하며 k-빈발 집합이 없을 경우 집합의 크기를 단계별로 줄여 나가며, Apriori 추천 모듈에서는 매장에서 구매한 제품 정보는 제외된다.

예를 들어 C_{11} 에 속한 고객이 S1 매장에 들어가서 제품을 구매하면, Apriori 추천 모듈에서 제품의 종류와 상관없이 S1 매장과 관련하여 동시에 방문하게 될 가능성이 높은 빈발 항목을 검색하게 된다. 만일 빈발 항목이 표 4와 같이 구성되어 있다면 S1을 포함하고 있는 가장 큰 빈발 집합, 즉 3-빈발집합에 포함된 $\{S1, S2, S3\}$, $\{S1, S2, S5\}$ 집합을 검색하여 고객의 단말기에 $\{S2, S3\}$, $\{S2, S5\}$ 를 추천하게 될 것이다. 이와 같은 과정은 고객의 구매가 발생할 때마다 계속해서 반복된다. 고객의 구매 활동을 추적하는 과정에서 S4와 같이 3-빈발집합에서 검색할 수 없는 매장을 고객이 방문했을 경우에는 한 단계 작은 빈발집합, 즉 2-빈발집합에서 추천할 매장을 검색하게 된다. 그러면 $\{S2, S4\}$ 집합이 검색될 것이며, 고객의 단말기에는 $\{S2\}$ 매장이 추천될 것이다.

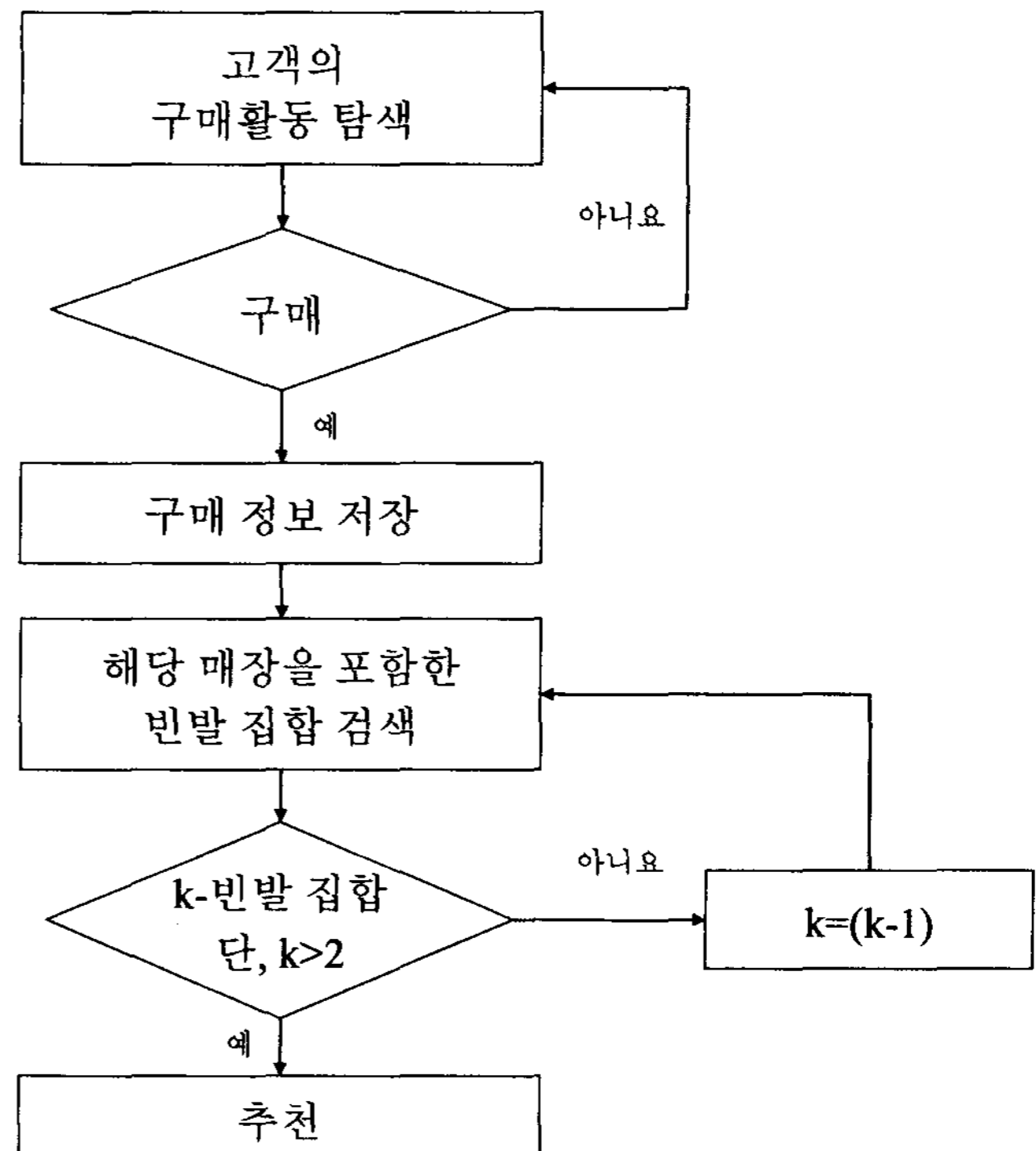


그림 5 - Apriori 추천 모듈 프로세스

5. 결론 및 향후 연구방향

지금까지의 추천시스템은 주로 제품 중심으로 논의되어 왔으나, 본 연구에서는 유틸리티 시장 환경에 적합한 매장 추천시스템을 제안하였다. 특히 기존의 매장 관련 추천시스템과는 다르게 단순한

매장 정보 제공보다는 고객의 선호도를 반영할 수 있는 추천시스템을 제안하였다. 본 연구에서 제안하고 있는 시스템은 가장 성공적인 추천 기법으로 알려져 있는 협업 필터링에 Apriori 알고리즘을 결합함으로써 매장을 추천 받고자 하는 고객에게 선호가 유사한 이웃들의 과거 구매행동을 통해 동시에 방문할 가능성이 높은 매장들을 집합으로 묶어서 추천해 준다.

그러나 이와 같은 추천서비스는 고객의 구매관련 정보만을 이용하여 날씨나 이벤트와 같은 상황정보 및 사용자의 다양한 요구를 반영하지 못하고 있다. 따라서 유비쿼터스 환경에서 보다 다이내믹한 고객 선호 반영을 위해서는 추가적인 연구가 필요할 것으로 판단된다. 또한 향후에는 본 연구에서 제안하고 있는 추천시스템의 성능을 검증하기 위해서 실제 데이터를 이용한 실험을 수행하도록 할 것이다.

Acknowledgement

이 연구는 서울시 산학연 협력사업 (과제번호 : 10802)의 재래시장 활성화를 위한 u-Market 개발 과제로부터 지원을 받아 수행되었다.

References

- [1] Breese, J., Heckerman, D., and Kadie, C. (1998). "Empirical analysis of predictive algorithms for collaborative filtering," *Proceedings of the UAI*.
- [2] Conner, M.O., and Herlocker, J. (1999). "Clustering Items for Collaborative Filtering," *Proceedings of the ACM SIGIR Workshop on Recommender Systems*, Berkeley, CA, August.
- [3] Davies, D.L., and Bouldin, D.W. (1979). "A cluster separation measure," *IEEE Transactions on Pattern Recognition and Machine Intelligence*, No.1, pp. 224-227.
- [4] Fano, A.E. (1998). "Shopper's Eye: Using Location-based Filtering for a Shopping Agent in the Physical World," *Proceedings of the Second International Conference on Autonomous Agents (Agents'98)*, pp. 416-421.
- [5] Hill, W., Stead, L., Rosenstein, M., Furnas, G. (1995). "Recommending and Evaluating Choices in a Virtual Community of Use," *Proceedings of CHI '95 Conference on Human Factors in Computing Systems*, pp.194~201.
- [6] Jin, X., Mobasher, B., and Zhou, Y. (2005). "A Web Recommendation System Based on Maximum Entropy," *Information Technology: Coding and Computing(ITCC)*, No. 1, pp. 213-218.
- [7] Keegan, S. and O'Hare, G.M.P, (2004). "Easishop: Delivering Cross Merchant Product Comparison Shopping for the Mobile User," *Proceedings of the 15th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC'04)*, 5th-8th September Barcelona, Spain, IEEE Press,.
- [8] Kohonen, T., (1990). "The Self-Organizing Map," *Proceedings of the IEEE*, Vol. 78, No. 9, pp. 1464-1480.
- [9] Konstan, J., Miller, B., Maltz, D., Herlocker, J., Gordon, L., and Riedl, J. (1997). "GroupLens: Applying Collaborative Filtering to Usenet News," *Communications of ACM*, Vol. 40, pp. 77-87.
- [10] Krulwich, B. (1997) "LIFESTYLE FINDER: Intelligent User Profiling Using Large-Scale Demographic Data," *Artificial Intelligence Magazine*, Vol. 18, No. 2, pp. 37-45.
- [11] Li, Y., Lu, L., and Xuefeng, L. (2005). "Hybrid Collaborative Filtering Method for Multiple-interests and Multiple-content Recommendation in E-Commerce," *Expert Systems with Applications*, Vol. 28, pp. 67-77.
- [12] Lin, W., Alvarez, S.A., and Ruiz, C. (2002). "Efficient Adaptive-Support Association Rule Mining for Recommender Systems," *Data Mining and Knowledge Discovery*, Vol. 6, pp. 83-105.
- [13] Montaner, M., Lopez, B., and Rosa, J.L. (2003). "A Taxonomy of Recommender Agents on the Internet," *Artificial Intelligence Review*, Vol. 19, pp. 285-330.
- [14] Park, D.J. (2003). "Co-evolution In uCommerce: Emerging Business Strategies and Technologies," *Telecommunicatins Review*, Vol. 13, No. 1, pp. 48~56.
- [15] Rensnick, P., Iacovou, N., Suchak, M., Bergstrom, P., and Riedl, J. (1994). "GroupLens: An Open Architecture for Collaborative Filtering of Netnews," *Proceedings of the Computer Supported Cooperative Work conference*, pp. 175~186.
- [16] Sarwar, B., Karypis, G., Konstan, J., and Riedl, J. (2000). "Analysis of Recommendation Algorithms for E-Commerce," *Proceedings of the ACM Conference on Electronic Commerce*, pp. 158-167.
- [17] Schafer, J.B., Konstan, J., and Riedl, J. (1999). "Recommender System in E - Commerce," *Proceedings of the ACM Conference on Electronic Commerce*.
- [18] Suryavanshi, B.S., Shiri, N., and Mudur, S.P. (2005). "Improving the Effectiveness of Model Based Recommender Systems for Highly Sparse and Noisy Web Usage Data," *Proceedings of the IEEE/WIC/MCM International Conference on Web Intelligence*.
- [19] Xue, G.R., Lin, C., Yang, Q., Xi, W., Zeng, H., Yu, Y., and Chen, Z. (2005). "Scalable Collaborative Filtering Using Cluster-based Smoothing," *Proceedings of the 2005 ACM SIGIR Conference*, pp. 114-121.
- [20] Han, J., and Kamber, M. (2001). "DataMining: Concepts and Techniques," Morgan Kaufmann.

[21] Laudon, K.C., and Laudon, J.P. (2007). "Essentials of Business Information Systems :seventh edition," Pearson Prentice Hall.